c246 Problem Set 8: Due April 22 at 5 PM to homework@c246.lbl.gov

1. (10 points) Describe a situation in which UPGMA method will fail.

2. (10 points) Draw an unrooted tree with four leaves. Indicate all possible placements of the root, and the rooted trees that correspond to each placement.

3. (25 points) Write a program that takes a protein alignment and computes a phylogenetic tree based on this alignment using UPGMA and Neighbor-Joining methods and a sensible distance score of your choice.

4. (25 points) Write a program that starts with a single sequences of 100 amino acids and simulates its evolution as follows:

- During each interval of time T, each amino acid of sequence I has a probability P(I) of mutating. If it mutates, choose a new amino acid randomly (ideally using a reasonable substitution matrix).
- At the end of every interval of time T, each sequence splits into two new sequences (i.e. there is a speciation event for every existing sequence). P(I) for the first of each pair of new sequences is equal to the P(I) for its parent sequence, the P(I) of the $2^{nd}$ sequence is equal to the P(I) of the parent sequence times a random number between M and 1/M (where M is a parameter of the program – note that when M = 1, where P(I)'s do not change). NOTE: The point of M is to have varying evolutionary rates along each branch.

Run the program

- A) 7 cycles with starting P(I) = 0.04 and M = 1
- B) 7 cycles with starting P(I) = 0.04 and M = 2

5. (30 points). For both A and B above, choose 10 random sequences and run the program from (3) using UPGMA and NJ. How do the computed trees (should be a total of 4) compare to the actual trees (from the program in (4) which generated the sequences).